

1 SHEPPARD MULLIN RICHTER & HAMPTON LLP  
2 Gary L. Halling, Cal. Bar No. 66087  
3 James L. McGinnis, Cal. Bar No. 95788  
4 Michael W. Scarborough, Cal. Bar No. 203524  
5 Mona Solouki, Cal. Bar No. 215145  
6 Four Embarcadero Center, 17th Floor  
7 San Francisco, CA 94111-4109  
8 Telephone: (415) 434-9100  
9 Facsimile: (415) 434-3947  
10 E-mail:[ghalling@sheppardmullin.com](mailto:ghalling@sheppardmullin.com)  
[jmcginnis@sheppardmullin.com](mailto:jmcginnis@sheppardmullin.com)  
[mscarborough@sheppardmullin.com](mailto:mscarborough@sheppardmullin.com)  
[monasolouki@sheppardmullin.com](mailto:monasolouki@sheppardmullin.com)

11 Attorneys for Defendants  
12 SAMSUNG SDI AMERICA, INC.,  
13 SAMSUNG SDI CO., LTD.,  
14 SAMSUNG SDI (MALAYSIA) SDN. BHD.,  
15 SAMSUNG SDI MEXICO S.A. DE C.V.,  
16 SAMSUNG SDI BRASIL LTDA.,  
17 SHENZEN SAMSUNG SDI CO., LTD. and  
18 TIANJIN SAMSUNG SDI CO., LTD.

19  
20 UNITED STATES DISTRICT COURT  
21 NORTHERN DISTRICT OF CALIFORNIA  
22  
23 (SAN FRANCISCO DIVISION)

24  
25 IN RE: CATHODE RAY TUBE (CRT)  
26 ANTITRUST LITIGATION

27 Case No. 3:07-md-05944 SC  
28 MDL No. 1917

29 This Document Relates to:

30 *CompuCom Systems, Inc. v. Hitachi, Ltd., et al.*,  
31 Case No. 3:11-cv-06396;

32 *Dell Inc., et al. v. Hitachi, Ltd., et al.*, Case  
33 No. 13-cv-02171;

34 *Sharp Electronics Corporation, Sharp*  
35 *Electronics Manufacturing Company of*  
36 *America, Inc. v. Hitachi, Ltd., et al.*, Case No.  
37 13-cv-1173;

38 *Sharp Electronics Corporation, Sharp*  
39 *Electronics Manufacturing Company of*  
40 *America, Inc. v. Koninklijke Philips Electronics*  
41 *N.V., et al.*, Case No. 13-cv-2776;

42  
43 **DECLARATION OF MONA**  
44 **SOLOUKI IN SUPPORT OF SDI**  
45 **DEFENDANTS' OPPOSITION TO**  
46 **CERTAIN PLAINTIFFS' MOTION**  
47 **TO PARTIALLY EXCLUDE**  
48 **EXPERT TESTIMONY OF DANIEL**  
49 **L. RUBINFELD**

*Electrograph Systems, Inc., et al. v. Hitachi, Ltd., et al.*, Case No. 3:11-cv-01656;

*Electrograph Systems, Inc., et al. v. Technicolor SA, et al.*, Case No. 3:13-cv-05724;

*Interbond Corp. of America v. Hitachi, Ltd., et al.*, Case No. 3:11-cv-06275;

*Interbond Corp. of America v. Technicolor SA, et al.*, Case No. 3:13-cv-05727;

*Office Depot, Inc. v. Hitachi, Ltd., et al.*, Case No. 3:11-cv-06276;

*Office Depot, Inc. v. Technicolor SA, et al.*, Case No. 3:13-cv-05726;

*P.C. Richard & Son Long Island Corp., et al. v. Hitachi, Ltd., et al.*, Case No. 3:12-cv-02648;

*P.C. Richard & Son Long Island Corp., et al. v. Technicolor SA, et al.*, Case No. 3:13-cv-05725;

*Sears, Roebuck & Co. and Kmart Corp. v. Technicolor SA.*, Case No. 3:13-cv-05262;

*Sears, Roebuck & Co. and Kmart Corp. v. Chunghwa Picture Tubes, Ltd., et al.*, Case No. 3:11-cv-05514;

*Siegel v. Hitachi, Ltd.*, Case No. 11-cv-05502;

*Siegel v. Technicolor SA.*, Case No. 13-cv-05261;

*Target Corp. v. Chunghwa Picture Tubes, Ltd., et al.*, Case No. 11-cv-05514;

*Target Corp. v. Technicolor SA*, Case No. 13-cv-05686;

*Tech Data Corp., et al. v. Hitachi, Ltd., et al.*, Case No. 3:13-cv-00157;

*Viewsonic Corporation v. Chunghwa Picture Tubes, Ltd., et al.*, Case No. 3:14-cv-002510;

REDACTED VERSION OF DOCUMENTS SOUGHT TO BE SEALED

1 I, Mona Solouki, declare as follows:

2       1. I am a special counsel at the law firm of Sheppard Mullin Richter & Hampton LLP,  
3 counsel of record for defendants Defendants Samsung SDI America, Inc.; Samsung SDI Co., Ltd.;  
4 Samsung SDI (Malaysia) SDN. Bhd.; Samsung SDI Mexico S.A. De C.V.; Samsung SDI Brasil  
5 Ltda.; Shenzhen Samsung SDI Co., Ltd.; and Tianjin Samsung SDI Co., Ltd. (collectively, "SDI").  
6 I submit this declaration in support of the SDI Defendants' Opposition to Certain Plaintiffs'  
7 Motion to Partially Exclude Testimony of Designated Expert Daniel L. Rubinfeld. I have personal  
8 knowledge of the facts set forth herein and, if called as a witness, I could and would competently  
9 testify thereto.

10      2. Attached hereto as Exhibit A is a true and correct copy of excerpts from the April  
11 15, 2014 expert report of Dr. Mohan Rao, Dell Plaintiffs' expert witness.

12      3. Attached hereto as Exhibit B is a true and correct copy of the April 15, 2014 expert  
13 report and exhibits of Kenneth G. Elzinga, expert witness for certain direct action plaintiffs  
14 ("DAPs").

15      4. Attached hereto as Exhibit C is a true and correct copy of excerpts from the April  
16 15, 2014 expert report of Janet S. Netz, expert witness for the indirect purchaser plaintiff ("IPP")  
17 class.

18      5. Attached hereto as Exhibit D is a true and correct copy of excerpts of Exhibit 1  
19 from the September 26, 2014 expert rebuttal reports of Alan S. Frankel, certain DAPs' expert  
20 witness.

21      6. Attached hereto as Exhibit E is a true and correct copy of excerpts from the July 3,  
22 2014 supplemental expert report of Jerry A. Hausman, Sharp Plaintiffs' expert witness.

23      7. Attached hereto as Exhibit F is a true and correct copy of excerpts from the  
24 September 26, 2014 rebuttal expert report of Jerry A. Hausman, Sharp Plaintiffs' expert witness.

25      8. Attached hereto as Exhibit G is a true and correct copy of the November 6, 2014  
26 sur-rebuttal expert report and exhibits of Daniel L. Rubinfeld, SDI's expert witness.

27      9. Attached hereto as Exhibit H is a true and correct copy of excerpts from the  
28 transcript of the July 17, 2014 deposition of Dr. Kenneth Elzinga, certain DAPs' expert witness.

10. Attached hereto as Exhibit I is a true and correct copy of the September 26, 2014 expert rebuttal report of Kenneth G. Elzinga, certain DAPs' expert witness.

11. Attached hereto as Exhibit J is a true and correct copy of the April 15, 2014 expert report of Jerry A. Hausman, Sharp Plaintiffs' expert witness.

12. Attached hereto as Exhibit K is a true and correct copy of excerpts from the transcript of the July 23, 2014 deposition of Dr. Jerry A. Hausman, Sharp Plaintiffs' expert witness.

13. Attached hereto as Exhibit L is a true and correct copy of excerpts from DAMODAR N. GUJARATI & DAWN C. PORTER, BASIC ECONOMETRICS (5th ed. McGraw Hill 2009).

14. Attached hereto as Exhibit M is a true and correct copy of excerpts from Daniel L. Rubinfeld, *Reference Guide on Multiple Regression*, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, at 310 (3d ed. 2011).

I declare under penalty of perjury under the laws of the United States of America  
that the foregoing is true and correct.

Executed this 16th day of January 2015 in San Francisco, California.

/s/ *Mona Solouki*  
Mona Solouki

# **Exhibit A**

## **[SUBMITTED UNDER SEAL]**

# **Exhibit B**

## **[SUBMITTED UNDER SEAL]**

# **Exhibit C**

## **[SUBMITTED UNDER SEAL]**

# **Exhibit D**

## **[SUBMITTED UNDER SEAL]**

# **Exhibit E**

## **[SUBMITTED UNDER SEAL]**

# **Exhibit F**

## **[SUBMITTED UNDER SEAL]**

# **Exhibit G**

## **[SUBMITTED UNDER SEAL]**

# **Exhibit H**

## **[SUBMITTED UNDER SEAL]**

# **Exhibit I**

## **[SUBMITTED UNDER SEAL]**

# **Exhibit J**

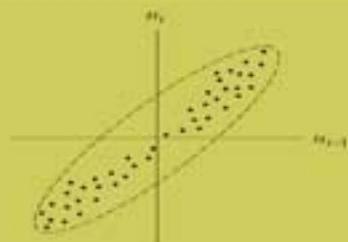
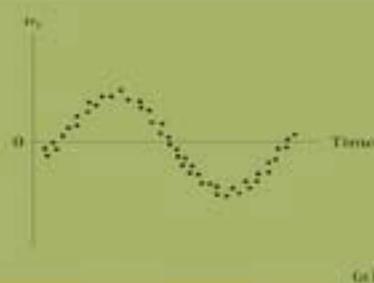
## **[SUBMITTED UNDER SEAL]**

# **Exhibit K**

## **[SUBMITTED UNDER SEAL]**

# Exhibit L

# BASIC ECONOMETRICS



Fifth Edition

Damodar N. Gujarati  
Dawn C. Porter

## 1.3 Statistical versus Deterministic Relationships

---

From the examples cited in Section 1.2, the reader will notice that in regression analysis we are concerned with what is known as the *statistical*, not *functional* or *deterministic*, dependence among variables, such as those of classical physics. In statistical relationships among variables we essentially deal with **random** or **stochastic**<sup>4</sup> variables, that is, variables that have probability distributions. In functional or deterministic dependency, on the other hand, we also deal with variables, but these variables are not random or stochastic.

The dependence of crop yield on temperature, rainfall, sunshine, and fertilizer, for example, is statistical in nature in the sense that the explanatory variables, although certainly important, will not enable the agronomist to predict crop yield exactly because of errors involved in measuring these variables as well as a host of other factors (variables) that collectively affect the yield but may be difficult to identify individually. Thus, there is bound to be some “intrinsic” or random variability in the dependent-variable crop yield that cannot be fully explained no matter how many explanatory variables we consider.

In deterministic phenomena, on the other hand, we deal with relationships of the type, say, exhibited by Newton’s law of gravity, which states: Every particle in the universe attracts every other particle with a force directly proportional to the product of their masses and inversely proportional to the square of the distance between them. Symbolically,  $F = k(m_1 m_2 / r^2)$ , where  $F$  = force,  $m_1$  and  $m_2$  are the masses of the two particles,  $r$  = distance, and  $k$  = constant of proportionality. Another example is Ohm’s law, which states: For metallic conductors over a limited range of temperature the current  $C$  is proportional to the voltage  $V$ ; that is,  $C = (\frac{1}{k})V$  where  $\frac{1}{k}$  is the constant of proportionality. Other examples of such deterministic relationships are Boyle’s gas law, Kirchhoff’s law of electricity, and Newton’s law of motion.

In this text we are not concerned with such deterministic relationships. Of course, if there are errors of measurement, say, in the  $k$  of Newton’s law of gravity, the otherwise deterministic relationship becomes a statistical relationship. In this situation, force can be predicted only approximately from the given value of  $k$  (and  $m_1$ ,  $m_2$ , and  $r$ ), which contains errors. The variable  $F$  in this case becomes a random variable.

## 1.4 Regression versus Causation

---

Although regression analysis deals with the dependence of one variable on other variables, it does not necessarily imply causation. In the words of Kendall and Stuart, “A statistical relationship, however strong and however suggestive, can never establish causal connection: our ideas of causation must come from outside statistics, ultimately from some theory or other.”<sup>5</sup>

<sup>4</sup>The word *stochastic* comes from the Greek word *stokhos* meaning “a bull’s eye.” The outcome of throwing darts on a dart board is a stochastic process, that is, a process fraught with misses.

<sup>5</sup>M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Charles Griffin Publishers, New York, vol. 2, 1961, chap. 26, p. 279.

In the crop-yield example cited previously, there is no *statistical reason* to assume that rainfall does not depend on crop yield. The fact that we treat crop yield as dependent on rainfall (among other things) is due to nonstatistical considerations: Common sense suggests that the relationship cannot be reversed, for we cannot control rainfall by varying crop yield.

In all the examples cited in Section 1.2 the point to note is that **a statistical relationship in itself cannot logically imply causation**. To ascribe causality, one must appeal to a priori or theoretical considerations. Thus, in the third example cited, one can invoke economic theory in saying that consumption expenditure depends on real income.<sup>6</sup>

## 1.5 Regression versus Correlation

---

Closely related to but conceptually very much different from regression analysis is **correlation analysis**, where the primary objective is to measure the *strength* or *degree* of *linear association* between two variables. The **correlation coefficient**, which we shall study in detail in Chapter 3, measures this strength of (linear) association. For example, we may be interested in finding the correlation (coefficient) between smoking and lung cancer, between scores on statistics and mathematics examinations, between high school grades and college grades, and so on. In regression analysis, as already noted, we are not primarily interested in such a measure. Instead, we try to estimate or predict the average value of one variable on the basis of the fixed values of other variables. Thus, we may want to know whether we can predict the average score on a statistics examination by knowing a student's score on a mathematics examination.

Regression and correlation have some fundamental differences that are worth mentioning. In regression analysis there is an asymmetry in the way the dependent and explanatory variables are treated. The dependent variable is assumed to be statistical, random, or stochastic, that is, to have a probability distribution. The explanatory variables, on the other hand, are assumed to have fixed values (in repeated sampling),<sup>7</sup> which was made explicit in the definition of regression given in Section 1.2. Thus, in Figure 1.2 we assumed that the variable age was fixed at given levels and height measurements were obtained at these levels. In correlation analysis, on the other hand, we treat any (two) variables symmetrically; there is no distinction between the dependent and explanatory variables. After all, the correlation between scores on mathematics and statistics examinations is the same as that between scores on statistics and mathematics examinations. Moreover, both variables are assumed to be random. As we shall see, most of the correlation theory is based on the assumption of randomness of variables, whereas most of the regression theory to be expounded in this book is conditional upon the assumption that the dependent variable is stochastic but the explanatory variables are fixed or nonstochastic.<sup>8</sup>

<sup>6</sup>But as we shall see in Chapter 3, classical regression analysis is based on the assumption that the model used in the analysis is the correct model. Therefore, the direction of causality may be implicit in the model postulated.

<sup>7</sup>It is crucial to note that the explanatory variables may be intrinsically stochastic, but for the purpose of regression analysis we assume that their values are fixed in repeated sampling (that is,  $X$  assumes the same values in various samples), thus rendering them in effect nonrandom or nonstochastic. But more on this in Chapter 3, Sec. 3.2.

<sup>8</sup>In advanced treatment of econometrics, one can relax the assumption that the explanatory variables are nonstochastic (see introduction to Part 2).

# **Exhibit M**

# Reference Manual on Scientific Evidence

*Third Edition*

Committee on the Development of the Third Edition of the  
Reference Manual on Scientific Evidence

Committee on Science, Technology, and Law  
Policy and Global Affairs

FEDERAL JUDICIAL CENTER

NATIONAL RESEARCH COUNCIL  
*OF THE NATIONAL ACADEMIES*

THE NATIONAL ACADEMIES PRESS  
Washington, D.C.  
**www.nap.edu**

*Reference Guide on Multiple Regression*

## I. Introduction and Overview

Multiple regression analysis is a statistical tool used to understand the relationship between or among two or more variables.<sup>1</sup> Multiple regression involves a variable to be explained—called the dependent variable—and additional explanatory variables that are thought to produce or be associated with changes in the dependent variable.<sup>2</sup> For example, a multiple regression analysis might estimate the effect of the number of years of work on salary. Salary would be the dependent variable to be explained; the years of experience would be the explanatory variable.

Multiple regression analysis is sometimes well suited to the analysis of data about competing theories for which there are several possible explanations for the relationships among a number of explanatory variables.<sup>3</sup> Multiple regression typically uses a single dependent variable and several explanatory variables to assess the statistical data pertinent to these theories. In a case alleging sex discrimination in salaries, for example, a multiple regression analysis would examine not only sex, but also other explanatory variables of interest, such as education and experience.<sup>4</sup> The employer-defendant might use multiple regression to argue that salary is a function of the employee's education and experience, and the employee-plaintiff might argue that salary is also a function of the individual's sex. Alternatively, in an antitrust cartel damages case, the plaintiff's expert might utilize multiple regression to evaluate the extent to which the price of a product increased during the period in which the cartel was effective, after accounting for costs and other variables unrelated to the cartel. The defendant's expert might use multiple

1. A variable is anything that can take on two or more values (e.g., the daily temperature in Chicago or the salaries of workers at a factory).

2. Explanatory variables in the context of a statistical study are sometimes called independent variables. See David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section II.A.1, in this manual. The guide also offers a brief discussion of multiple regression analysis. *Id.*, Section V.

3. Multiple regression is one type of statistical analysis involving several variables. Other types include matching analysis, stratification, analysis of variance, probit analysis, logit analysis, discriminant analysis, and factor analysis.

4. Thus, in *Ottaviani v. State University of New York*, 875 F.2d 365, 367 (2d Cir. 1989) (citations omitted), *cert. denied*, 493 U.S. 1021 (1990), the court stated:

In disparate treatment cases involving claims of gender discrimination, plaintiffs typically use multiple regression analysis to isolate the influence of gender on employment decisions relating to a particular job or job benefit, such as salary.

The first step in such a regression analysis is to specify all of the possible “legitimate” (i.e., non-discriminatory) factors that are likely to significantly affect the dependent variable and which could account for disparities in the treatment of male and female employees. By identifying those legitimate criteria that affect the decisionmaking process, individual plaintiffs can make predictions about what job or job benefits similarly situated employees should ideally receive, and then can measure the difference between the predicted treatment and the actual treatment of those employees. If there is a disparity between the predicted and actual outcomes for female employees, plaintiffs in a disparate treatment case can argue that the net “residual” difference represents the unlawful effect of discriminatory animus on the allocation of jobs or job benefits.

*Reference Manual on Scientific Evidence*

regression to suggest that the plaintiff's expert had omitted a number of price-determining variables.

More generally, multiple regression may be useful (1) in determining whether a particular effect is present; (2) in measuring the magnitude of a particular effect; and (3) in forecasting what a particular effect would be, but for an intervening event. In a patent infringement case, for example, a multiple regression analysis could be used to determine (1) whether the behavior of the alleged infringer affected the price of the patented product, (2) the size of the effect, and (3) what the price of the product would have been had the alleged infringement not occurred.

Over the past several decades, the use of multiple regression analysis in court has grown widely. Regression analysis has been used most frequently in cases of sex and race discrimination<sup>5</sup> antitrust violations,<sup>6</sup> and cases involving class cer-

5. Discrimination cases using multiple regression analysis are legion. *See, e.g.*, Bazemore v. Friday, 478 U.S. 385 (1986), *on remand*, 848 F.2d 476 (4th Cir. 1988); Csicseri v. Bowsher, 862 F. Supp. 547 (D.D.C. 1994) (age discrimination), *aff'd*, 67 F.3d 972 (D.C. Cir. 1995); EEOC v. General Tel. Co., 885 F.2d 575 (9th Cir. 1989), *cert. denied*, 498 U.S. 950 (1990); Bridgeport Guardians, Inc. v. City of Bridgeport, 735 F. Supp. 1126 (D. Conn. 1990), *aff'd*, 933 F.2d 1140 (2d Cir.), *cert. denied*, 502 U.S. 924 (1991); Bickerstaff v. Vassar College, 196 F.3d 435, 448-49 (2d Cir. 1999) (sex discrimination); McReynolds v. Sodexho Marriott, 349 F. Supp. 2d 1 (D.C. Cir. 2004) (race discrimination); Hnot v. Willis Group Holdings Ltd., 228 F.R.D. 476 (S.D.N.Y. 2005) (gender discrimination); Carpenter v. Boeing Co., 456 F.3d 1183 (10th Cir. 2006) (sex discrimination); Coward v. ADT Security Systems, Inc., 140 F.3d 271, 274-75 (D.C. Cir. 1998); Smith v. Virginia Commonwealth Univ., 84 F.3d 672 (4th Cir. 1996) (en banc); Hemmings v. Tidyman's Inc., 285 F.3d 1174, 1184-86 (9th Cir. 2000); Mehus v. Emporia State University, 222 F.R.D. 455 (D. Kan. 2004) (sex discrimination); Gutierrez v. Johnson & Johnson, 2006 WL 3246605 (D.N.J. Nov. 6, 2006) (race discrimination); Morgan v. United Parcel Service, 380 F.3d 459 (8th Cir. 2004) (racial discrimination). *See also* Keith N. Hylton & Vincent D. Rougeau, *Lending Discrimination: Economic Theory, Econometric Evidence, and the Community Reinvestment Act*, 85 Geo. L.J. 237, 238 (1996) ("regression analysis is probably the best empirical tool for uncovering discrimination").

6. *E.g.*, United States v. Brown Univ., 805 F. Supp. 288 (E.D. Pa. 1992) (price fixing of college scholarships), *rev'd*, 5 F.3d 658 (3d Cir. 1993); Petrucci's IGA Supermarkets, Inc. v. Darling-Delaware Co., 998 F.2d 1224 (3d Cir.), *cert. denied*, 510 U.S. 994 (1993); Ohio v. Louis Trauth Dairy, Inc., 925 F. Supp. 1247 (S.D. Ohio 1996); *In re Chicken Antitrust Litig.*, 560 F. Supp. 963, 993 (N.D. Ga. 1980); New York v. Kraft Gen. Foods, Inc., 926 F. Supp. 321 (S.D.N.Y. 1995); Freeland v. AT&T, 238 F.R.D. 130 (S.D.N.Y. 2006); *In re Pressure Sensitive Labelstock Antitrust Litig.*, 2007 U.S. Dist. LEXIS 85466 (M.D. Pa. Nov. 19, 2007); *In re Linerboard Antitrust Litig.*, 497 F. Supp. 2d 666 (E.D. Pa. 2007) (price fixing by manufacturers of corrugated boards and boxes); *In re Polypropylene Carpet Antitrust Litig.*, 93 F. Supp. 2d 1348 (N.D. Ga. 2000); *In re OSB Antitrust Litig.*, 2007 WL 2253418 (E.D. Pa. Aug. 3, 2007) (price fixing of Oriented Strand Board, also known as "waferboard"); *In re TFT-LCD (Flat Panel) Antitrust Litig.*, 267 F.R.D. 583 (N.D. Cal. 2010).

For a broad overview of the use of regression methods in antitrust, see ABA Antitrust Section, *Econometrics: Legal, Practical and Technical Issues* (John Harkrider & Daniel Rubinfeld, eds. 2005). *See also* Jerry Hausman et al., *Competitive Analysis with Differentiated Products*, 34 Annales D'Économie et de Statistique 159 (1994); Gregory J. Werden, *Simulating the Effects of Differentiated Products Mergers: A Practical Alternative to Structural Merger Policy*, 5 Geo. Mason L. Rev. 363 (1997).

*Reference Guide on Multiple Regression*

tification (under Rule 23).<sup>7</sup> However, there are a range of other applications, including census undercounts,<sup>8</sup> voting rights,<sup>9</sup> the study of the deterrent effect of the death penalty,<sup>10</sup> rate regulation,<sup>11</sup> and intellectual property.<sup>12</sup>

7. In antitrust, the circuits are currently split as to the extent to which plaintiffs must prove that common elements predominate over individual elements. *E.g.*, compare *In Re Hydrogen Peroxide Litig.*, 522 F.2d 305 (3d Cir. 2008) with *In Re Cardizem CD Antitrust Litig.*, 391 F.3d 812 (6th Cir. 2004). For a discussion of use of multiple regression in evaluating class certification, see Bret M. Dickey & Daniel L. Rubinfeld, *Antitrust Class Certification: Towards an Economic Framework*, 66 N.Y.U. Ann. Surv. Am. L. 459 (2010) and John H. Johnson & Gregory K. Leonard, *Economics and the Rigorous Analysis of Class Certification in Antitrust Cases*, 3 J. Competition L. & Econ. 341 (2007).

8. See, e.g., *City of New York v. U.S. Dep't of Commerce*, 822 F. Supp. 906 (E.D.N.Y. 1993) (decision of Secretary of Commerce not to adjust the 1990 census was not arbitrary and capricious), vacated, 34 F.3d 1114 (2d Cir. 1994) (applying heightened scrutiny), *rev'd sub nom. Wisconsin v. City of New York*, 517 U.S. 565 (1996); *Carey v. Klutznick*, 508 F. Supp. 420, 432–33 (S.D.N.Y. 1980) (use of reasonable and scientifically valid statistical survey or sampling procedures to adjust census figures for the differential undercount is constitutionally permissible), stay granted, 449 U.S. 1068 (1980), *rev'd on other grounds*, 653 F.2d 732 (2d Cir. 1981), cert. denied, 455 U.S. 999 (1982); *Young v. Klutznick*, 497 F. Supp. 1318, 1331 (E.D. Mich. 1980), *rev'd on other grounds*, 652 F.2d 617 (6th Cir. 1981), cert. denied, 455 U.S. 939 (1982).

9. Multiple regression analysis was used in suits charging that at-large areawide voting was instituted to neutralize black voting strength, in violation of section 2 of the Voting Rights Act, 42 U.S.C. § 1973 (1988). Multiple regression demonstrated that the race of the candidates and that of the electorate were determinants of voting. See *Williams v. Brown*, 446 U.S. 236 (1980); *Rodriguez v. Pataki*, 308 F. Supp. 2d 346, 414 (S.D.N.Y. 2004); *United States v. Vill. of Port Chester*, 2008 U.S. Dist. LEXIS 4914 (S.D.N.Y. Jan. 17, 2008); *Meza v. Galvin*, 322 F. Supp. 2d 52 (D. Mass. 2004) (violation of VRA with regard to Hispanic voters in Boston); *Bone Shirt v. Hazeltine*, 336 F. Supp. 2d 976 (D.S.D. 2004) (violations of VRA with regard to Native American voters in South Dakota); *Georgia v. Ashcroft*, 195 F. Supp. 2d 25 (D.D.C. 2002) (redistricting of Georgia's state and federal legislative districts); *Benavidez v. City of Irving*, 638 F. Supp. 2d 709 (N.D. Tex. 2009) (challenge of city's at-large voting scheme). For commentary on statistical issues in voting rights cases, see, e.g., *Statistical and Demographic Issues Underlying Voting Rights Cases*, 15 Evaluation Rev. 659 (1991); Stephen P. Klein et al., *Ecological Regression Versus the Secret Ballot*, 31 Jurimetrics J. 393 (1991); James W. Loewen & Bernard Grofman, *Recent Developments in Methods Used in Vote Dilution Litigation*, 21 Urb. Law. 589 (1989); Arthur Lupia & Kenneth McCue, *Why the 1980s Measures of Racially Polarized Voting Are Inadequate for the 1990s*, 12 Law & Pol'y 353 (1990).

10. See, e.g., *Gregg v. Georgia*, 428 U.S. 153, 184–86 (1976). For critiques of the validity of the deterrence analysis, see National Research Council, *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates* (Alfred Blumstein et al. eds., 1978); Richard O. Lempert, *Desert and Deterrence: An Assessment of the Moral Bases of the Case for Capital Punishment*, 79 Mich. L. Rev. 1177 (1981); Hans Zeisel, *The Deterrent Effect of the Death Penalty: Facts v. Faith*, 1976 Sup. Ct. Rev. 317; and John Donohue & Justin Wolfers, *Uses and Abuses of Statistical Evidence in the Death Penalty Debate*, 58 Stan. L. Rev. 787 (2005).

11. See, e.g., *Time Warner Entertainment Co. v. FCC*, 56 F.3d 151 (D.C. Cir. 1995) (challenge to FCC's application of multiple regression analysis to set cable rates), cert. denied, 516 U.S. 1112 (1996); *Appalachian Power Co. v. EPA*, 135 F.3d 791 (D.C. Cir. 1998) (challenging the EPA's application of regression analysis to set nitrous oxide emission limits); *Consumers Util. Rate Advocacy Div. v. Ark. PSC*, 99 Ark. App. 228 (Ark. Ct. App. 2007) (challenging an increase in nongas rates).

12. See *Polaroid Corp. v. Eastman Kodak Co.*, No. 76-1634-MA, 1990 WL 324105, at \*29, \*62–63 (D. Mass. Oct. 12, 1990) (damages awarded because of patent infringement), amended by No.

*Reference Manual on Scientific Evidence*

Multiple regression analysis can be a source of valuable scientific testimony in litigation. However, when inappropriately used, regression analysis can confuse important issues while having little, if any, probative value. In *EEOC v. Sears, Roebuck & Co.*,<sup>13</sup> in which Sears was charged with discrimination against women in hiring practices, the Seventh Circuit acknowledged that “[m]ultiple regression analyses, designed to determine the effect of several independent variables on a dependent variable, which in this case is hiring, are an accepted and common method of proving disparate treatment claims.”<sup>14</sup> However, the court affirmed the district court’s findings that the “E.E.O.C.’s regression analyses did not ‘accurately reflect Sears’ complex, nondiscriminatory decision-making processes’” and that the “‘E.E.O.C.’s statistical analyses [were] so flawed that they lack[ed] any persuasive value.’”<sup>15</sup> Serious questions also have been raised about the use of multiple regression analysis in census undercount cases and in death penalty cases.<sup>16</sup>

The Supreme Court’s rulings in *Daubert* and *Kumho Tire* have encouraged parties to raise questions about the admissibility of multiple regression analyses.<sup>17</sup> Because multiple regression is a well-accepted scientific methodology, courts have frequently admitted testimony based on multiple regression studies, in some cases over the strong objection of one of the parties.<sup>18</sup> However, on some occasions courts have excluded expert testimony because of a failure to utilize a multiple regression methodology.<sup>19</sup> On other occasions, courts have rejected regression

76-1634-MA, 1991 WL 4087 (D. Mass. Jan. 11, 1991); Estate of Vane v. The Fair, Inc., 849 F.2d 186, 188 (5th Cir. 1988) (lost profits were the result of copyright infringement), *cert. denied*, 488 U.S. 1008 (1989); Louis Vuitton Malletier v. Dooney & Bourke, Inc., 525 F. Supp. 2d 576, 664 (S.D.N.Y. 2007) (trademark infringement and unfair competition suit). The use of multiple regression analysis to estimate damages has been contemplated in a wide variety of contexts. See, e.g., David Baldus et al., *Improving Judicial Oversight of Jury Damages Assessments: A Proposal for the Comparative Additur/Remittitur Review of Awards for Nonpecuniary Harms and Punitive Damages*, 80 Iowa L. Rev. 1109 (1995); Talcott J. Franklin, *Calculating Damages for Loss of Parental Nurture Through Multiple Regression Analysis*, 52 Wash. & Lee L. Rev. 271 (1997); Roger D. Blair & Amanda Kay Esquibel, *Yardstick Damages in Lost Profit Cases: An Econometric Approach*, 72 Denv. U. L. Rev. 113 (1994). Daniel Rubinfeld, *Quantitative Methods in Antitrust*, in 1 *Issues in Competition Law and Policy* 723 (2008).

13. 839 F.2d 302 (7th Cir. 1988).

14. *Id.* at 324 n.22.

15. *Id.* at 348, 351 (quoting *EEOC v. Sears, Roebuck & Co.*, 628 F. Supp. 1264, 1342, 1352 (N.D. Ill. 1986)). The district court commented specifically on the “severe limits of regression analysis in evaluating complex decision-making processes.” 628 F. Supp. at 1350.

16. See David H. Kaye & David A. Freedman, Reference Guide on Statistics, Sections II.A.3, B.1, in this manual.

17. *Daubert v. Merrill Dow Pharmas., Inc.* 509 U.S. 579 (1993); *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 147 (1999) (expanding the *Daubert*’s application to nonscientific expert testimony).

18. See *Newport Ltd. v. Sears, Roebuck & Co.*, 1995 U.S. Dist. LEXIS 7652 (E.D. La. May 26, 1995). See also Petruzzii’s *IGA Supermarkets*, *supra* note 6, 998 F.2d at 1240, 1247 (finding that the district court abused its discretion in excluding multiple regression-based testimony and reversing the grant of summary judgment to two defendants).

19. See, e.g., *In re Executive Telecard Ltd. Sec. Litig.*, 979 F. Supp. 1021 (S.D.N.Y. 1997).

*Reference Guide on Multiple Regression*

studies that did not have an adequate foundation or research design with respect to the issues at hand.<sup>20</sup>

In interpreting the results of a multiple regression analysis, it is important to distinguish between correlation and causality. Two variables are correlated—that is, associated with each other—when the events associated with the variables occur more frequently together than one would expect by chance. For example, if higher salaries are associated with a greater number of years of work experience, and lower salaries are associated with fewer years of experience, there is a positive correlation between salary and number of years of work experience. However, if higher salaries are associated with less experience, and lower salaries are associated with more experience, there is a negative correlation between the two variables.

A correlation between two variables does not imply that one event causes the second. Therefore, in making causal inferences, it is important to avoid *spurious correlation*.<sup>21</sup> Spurious correlation arises when two variables are closely related but bear no causal relationship because they are both caused by a third, unexamined variable. For example, there might be a negative correlation between the age of certain skilled employees of a computer company and their salaries. One should not conclude from this correlation that the employer has necessarily discriminated against the employees on the basis of their age. A third, unexamined variable, such as the level of the employees' technological skills, could explain differences in productivity and, consequently, differences in salary.<sup>22</sup> Or, consider a patent infringement case in which increased sales of an allegedly infringing product are associated with a lower price of the patented product.<sup>23</sup> This correlation would be spurious if the two products have their own noncompetitive market niches and the lower price is the result of a decline in the production costs of the patented product.

Pointing to the possibility of a spurious correlation will typically not be enough to dispose of a statistical argument. It may be appropriate to give little weight to such an argument absent a showing that the correlation is relevant. For example, a statistical showing of a relationship between technological skills

20. See *City of Tuscaloosa v. Harcros Chemicals, Inc.*, 158 F.2d 548 (11th Cir. 1998), in which the court ruled plaintiffs' regression-based expert testimony inadmissible and granted summary judgment to the defendants. See also *American Booksellers Ass'n v. Barnes & Noble, Inc.*, 135 F. Supp. 2d 1031, 1041 (N.D. Cal. 2001), in which a model was said to contain "too many assumptions and simplifications that are not supported by real-world evidence," and *Obrey v. Johnson*, 400 F.3d 691 (9th Cir. 2005).

21. See David H. Kaye & David A. Freedman, Reference Guide on Statistics, Section V.B.3, in this manual.

22. See, e.g., *Sheehan v. Daily Racing Form Inc.*, 104 F.3d 940, 942 (7th Cir.) (rejecting plaintiff's age discrimination claim because statistical study showing correlation between age and retention ignored the "more than remote possibility that age was correlated with a legitimate job-related qualification"), cert. denied, 521 U.S. 1104 (1997).

23. In some particular cases, there are statistical tests that allow one to reject claims of causality. For a brief description of these tests, which were developed by Jerry Hausman, see Robert S. Pindyck & Daniel L. Rubinfeld, *Econometric Models and Economic Forecasts* § 7.5 (4th ed. 1997).

*Reference Manual on Scientific Evidence*

and worker productivity might be required in the age discrimination example, above.<sup>24</sup>

Causality cannot be inferred by data analysis alone; rather, one must infer that a causal relationship exists on the basis of an underlying causal theory that explains the relationship between the two variables. Even when an appropriate theory has been identified, causality can never be inferred directly. One must also look for empirical evidence that there is a causal relationship. Conversely, the fact that two variables are correlated does not guarantee the existence of a relationship; it could be that the model—a characterization of the underlying causal theory—does not reflect the correct interplay among the explanatory variables. In fact, the absence of correlation does not guarantee that a causal relationship does not exist. Lack of correlation could occur if (1) there are insufficient data, (2) the data are measured inaccurately, (3) the data do not allow multiple causal relationships to be sorted out, or (4) the model is specified wrongly because of the omission of a variable or variables that are related to the variable of interest.

There is a tension between any attempt to reach conclusions with near certainty and the inherently uncertain nature of multiple regression analysis. In general, the statistical analysis associated with multiple regression allows for the expression of uncertainty in terms of probabilities. The reality that statistical analysis generates probabilities concerning relationships rather than certainty should not be seen in itself as an argument against the use of statistical evidence, or worse, as a reason to not admit that there is uncertainty at all. The only alternative might be to use less reliable anecdotal evidence.

This reference guide addresses a number of procedural and methodological issues that are relevant in considering the admissibility of, and weight to be accorded to, the findings of multiple regression analyses. It also suggests some standards of reporting and analysis that an expert presenting multiple regression analyses might be expected to meet. Section II discusses research design—how the multiple regression framework can be used to sort out alternative theories about a case. The guide discusses the importance of choosing the appropriate specification of the multiple regression model and raises the issue of whether multiple regression is appropriate for the case at issue. Section III accepts the regression framework and concentrates on the interpretation of the multiple regression results from both a statistical and a practical point of view. It emphasizes the distinction between regression results that are statistically significant and results that are meaningful to the trier of fact. It also points to the importance of evaluating the robustness

<sup>24</sup>. See, e.g., *Allen v. Seidman*, 881 F.2d 375 (7th Cir. 1989) (judicial skepticism was raised when the defendant did not submit a logistic regression incorporating an omitted variable—the possession of a higher degree or special education; defendant's attack on statistical comparisons must also include an analysis that demonstrates that comparisons are flawed). The appropriate requirements for the defendant's showing of spurious correlation could, in general, depend on the discovery process. See, e.g., *Boykin v. Georgia Pac. Co.*, 706 F.2d 1384 (1983) (criticism of a plaintiff's analysis for not including omitted factors, when plaintiff considered all information on an application form, was inadequate).

*Reference Guide on Multiple Regression*

of regression analyses, i.e., seeing the extent to which the results are sensitive to changes in the underlying assumptions of the regression model. Section IV briefly discusses the qualifications of experts and suggests a potentially useful role for court-appointed neutral experts. Section V emphasizes procedural aspects associated with use of the data underlying regression analyses. It encourages greater pretrial efforts by the parties to attempt to resolve disputes over statistical studies.

Throughout the main body of this guide, hypothetical examples are used as illustrations. Moreover, the basic “mathematics” of multiple regression has been kept to a bare minimum. To achieve that goal, the more formal description of the multiple regression framework has been placed in the Appendix. The Appendix is self-contained and can be read before or after the text. The Appendix also includes further details with respect to the examples used in the body of this guide.

## II. Research Design: Model Specification

Multiple regression allows the testifying economist or other expert to choose among alternative theories or hypotheses and assists the expert in distinguishing correlations between variables that are plainly spurious from those that may reflect valid relationships.

### *A. What Is the Specific Question That Is Under Investigation by the Expert?*

Research begins with a clear formulation of a research question. The data to be collected and analyzed must relate directly to this question; otherwise, appropriate inferences cannot be drawn from the statistical analysis. For example, if the question at issue in a patent infringement case is what price the plaintiff's product would have been but for the sale of the defendant's infringing product, sufficient data must be available to allow the expert to account statistically for the important factors that determine the price of the product.

### *B. What Model Should Be Used to Evaluate the Question at Issue?*

Model specification involves several steps, each of which is fundamental to the success of the research effort. Ideally, a multiple regression analysis builds on a theory that describes the variables to be included in the study. A typical regression model will include one or more dependent variables, each of which is believed to be causally related to a series of explanatory variables. Because we cannot be certain that the explanatory variables are themselves unaffected or independent of the influence of the dependent variable (at least at the point of initial study), the explanatory